CLAIMS

What is claimed is:

1. A method for creating a database for managing multiple types of biological information comprising:

5   obtaining a form of biological information,

inputting the biological information into the database as a new record, wherein the record is associated with a unique identifier,

comparing the information in the record to the information already present in the database,

10   determining whether the information in the new record already exists in the database,

adding the information to the database if it is not redundant to the database information, thereby forming a set of records in the database, where each record is associated with a unique identifier,

15   creating at least one module for a specific type of biological information that is associated with each unique identifier,

obtaining a form of biological information associated with a module in the database,

associating the biological information with the correct module in the database,

20   and associating the biological information with the correct unique identifier.

2. A method of creating an executive summary of biologically significant information, comprising:

inputting biologically significant information in a database;

checking the biologically significant information against the database for

25   redundancy;

sending sequences from the biologically significant information to a second database for comparison;

receiving replies from the second database in response to a comparison query;

saving the replies in the database, thereby creating a module;

collecting all of the modules associated with each identifier; and

outputting the information contained in the modules for each unique identifier in an executive summary.

3. A method of displaying an executive summary of biologically significant information on a computer wherein the computer comprises a processing means, a memory means, an input means and an output means comprising:

collecting information from individual information modules related to a unique identifier, wherein the unique identifier number is associated with a particular record;

producing a coordinated display of information from the individual modules; and

displaying the information from the individual modules using a visual display means producing the executive summary.

4. A method of displaying an executive summary containing information related to a unique identifier associated with a first set of sequences comprising:

(a) determining a first set of sequences;

(b) providing a computer system having a memory means, a data input means, and a visual display means, the memory means containing the first set of sequences, and modules containing information to be coordinated with the first set of sequences, and the memory means being operable to retrieve coordinate data from the memory means and to display an executive summary on the visual display means, the executive summary containing a representation of the first set of sequences, and information from the modules;

(c) uploading information from a second database containing sequence comparison data to the computer system;

(d) creating a module based on information obtained from the second database containing sequence comparison data;

(e) searching for other modules associated with the unique

identifier;

    (f) creating an executive summary containing information from the modules;

    (g) displaying the executive summary containing information on the first set of sequences and all the modules associated therewith.

5. A method of comparing a first set of sequences to a second set of sequences, the method comprising:

a) uploading the first set of sequences associated with a unique identifier contained as a module in a record in a first database into a network switch node,

b) uploading the second set of sequences contained in a second database into the network switch node,

c) parsing the first set of sequences into subsets of sequences,

d) allocating each subset of sequences to a search node,

e) downloading the second set of sequences to each search node,

f) comparing the subset of sequences to the second set of sequences on the search node, thereby forming an alignment, or comparison, of the first set of sequences and the second set of sequences,

g) monitoring the status of each comparison on each search node, until a particular search node completes the comparison of the subset of sequences being performed, thereby forming a completed node,

h) identifying the sequences in the subset of sequence on each node other than the completed node that have not yet been compared to the second set of sequences forming a set of remaining sequences,

i) parsing the set of remaining sequences into a second subset of sequences,

j) allocating the second subset of sequences onto each node, and

k) comparing the second subset of sequences to the second set of sequences,

l) and repeating steps g-k until each sequence in the first set of sequences has been compared to each sequence in the second set of sequences,

m) updating the information in the first database with the results of the

comparison of the first set of sequences to the second set of sequences

6. The method of claim 5, wherein the network database is in communication with a set of computer nodes and is scalable.

7. The method of claim 5, wherein there are three or more search nodes.

8. The method of claim 5, wherein the databases use Cold Fusion.

9. The method of claim 5, wherein the databases use Oracle.

10. The method of claim 5, further comprising the step of identifying which records in the first database have changed after the step of updating.

11. The method of claim 10, wherein a report is generated that indicates which records have changed since the last updating.

12. The method of claim 11, wherein the report is automatically sent via e-mail to a predetermined address.

13. The method of claim 10, wherein the changed information is flagged, the flags being searchable in the database.

14. The method of claim 5, wherein the second database is a mirror database.

15. The method of claim 14, wherein the mirror database mines a National Center for Biotechnology Information database.

16. The method of claim 14, wherein the mirror database mines Genbank, Pfam, Prodom, Prosite, Tmpred and Signal P database.

17. The method of claim 15, wherein the mirror database mines GenBank.

18. The method of claim 5, wherein the search node performs a BLAST search.

19. The method of claim 5, wherein the first database is SMEDDb.

20. The method of claim 5, wherein the second database comprises an HTML file readable by a web browser.

21. The method of claim 20, wherein the HTML file incorporates an image relating to a sequence.

22. The method of claim 20, further comprising accessing the HTML file remotely through a computer network.

23. The method of claim 5, wherein the various modules can be viewed in the executive summary.

24. The method of claim 5, wherein at least one module comprises the biologically significant information itself.

25. The method of claim 24, wherein the biologically significant information comprises sequence data.

26. The method of claim 5, wherein the first set of sequence data comprise cDNA data.

27. The method of claim 5, wherein the first set of sequences comprise expressed sequence tags.

28. The method of claim 5, wherein a module comprises gene expression patterns.

29. The method of claim 5, wherein a module comprises sequence comparison data obtained from the second database.

30. The method of claim 5, wherein a module comprises hybridization data.

31. The method of claim 30, wherein the module comprises *in situ* hydribidation data.

32. The method of claim 30, wherein the module comprises two hybrid data.

33. The method of claim 5, wherein a module comprises pharmacology data.

34. The method of claim 5, wherein a module comprises immunohistological data.

35. The method of claim 5, wherein a module comprises expression patterns.

36. The method of claim 5, wherein the module comprises information from a publicly accessible database.

37. The method of claim 5, further comprising analyzing the sequence comparison data to determine categories, subcategories and keywords for the unique identification number.

38. The method of claim 5, wherein the biologically significant information can be sorted by any of the characteristics associated with the modules.

39. The method of claim 38, wherein the information in the modules is associated with an executive summary.

40. The method of claim 5, wherein the second database is continually updated on a separate node.

41. The method of claim 40, wherein the updating of the second database occurs over the Internet.

42. The method of claim 5, wherein the first database comprises a module for spatial information and a module for temporal information.

43. The method of claim 5, further comprising providing a search interface accessible by a web browser.

44. A computer system for comparing a first set of sequences to a second set of sequences, the system comprising a first database containing a first set of sequences, a second database containing a second set of sequences, a network switch in communication with both the first and second databases.

45. The system of claim 44, wherein the network switch is also in communication with a set of computer search nodes.

46. The system of claim 44, wherein the system is scalable.

47. The system of claim 44, wherein there are two or more computer nodes.

48. The system of claim 44, wherein the databases use Cold Fusion.

49. The system of claim 44, wherein the databases use Oracle.

50. The system of claim 44, wherein the second database is a mirror database.

51. The system of claim 50, wherein the mirror database mines Genbank, Pfam, Prodom, Prosite, Tmpred and Signal P data.

52. The system of claim 50, wherein the mirror database mines a National Center of Biotechnology Information center database.

53. The system of claim 52, wherein the mirror database mines GenBank.

54. The system of claim 44, wherein the first database is SMEDDb.

55. The system of claim 44, wherein the database comprises an HTML file readable by a web browser.

56. The system of claim 55, wherein the HTML file incorporates an image relating to

a sequence.

57. The system of claim 55, wherein the HTML file can be accessed remotely through a computer network.

58. The system of claim 44, wherein a file and services server can be used to access the network.

59. The system of claim 44, wherein the network is the Internet.

60. The system of claim 44, wherein the network uses FTP.

61. The system of claim 44, wherein the database comprises spatial information and temporal information.

62. The system of claim 44, further comprising a search interface accessible by a web browser.

63. A computer system having a memory means, a data input means, and a visual display means, the memory means containing the first set of sequences, and modules containing information to be coordinated with the first set of sequences, and the memory means being operable to retrieve coordinate data from the memory means and to display an executive summary on the visual display means, the executive summary containing a representation of the first set of sequences, and information from the modules.

64. A computer system comprising a cluster computer, wherein the system can semi-automatically process a plurality of Blast searches when the databases change, producing a dynamic database that is regularly and automatically updated.

65. A computer cluster comprising,

a first database node, a second database node, a network switch and at least two computer search nodes,

wherein the network switch is in communication with the first database node, the second database node, and the computer search nodes.

66. The system of claim 65,

wherein the first database node comprises a database of biologically significant information.

67. The system of claim 65,

    wherein the second database node comprises a database that is mirrored.

68. The system of claim 65,

    wherein the network switch uploads the information from the first database and uploads the information from the second database.

69. The system of claim 65,

    wherein the network switch parses the information from the first database into a number of subsets equal to the number of computer search nodes and distributes one subset to each computer search node.

70. The system of claim 69,

    wherein the network switch downloads the second database to each computer search node.

71. The system of claim 70,

    wherein the network switch monitors the activity on the computer search nodes and when the activity on one computer search node is complete identifies the activity remaining to be completed on the other computer search nodes, parses the remaining activity into a second set of subsets equal to the number of computer search nodes in the system, and distributes one second subset to each computer search node.

72. The system of claim 65,

    wherein the second database node is continually updated.

73. The system of claim 65, wherein the first database has at least 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 3, 5, 8, 10, 12, 15, 20, 30, 40, 50, 75, or 100 gigabytes of data.

74. The system of claim 65, wherein the periodic search is performed at least three times, producing a first, second, and third generation of the periodic output.

75. The system of claim 74, wherein the first and second generations second period search producesdatabase is stored in a

76. The system of claim 65, wherein the second database is a dynamic database, in

which there are at least a first generation, a second generation, and a third generation of the database.

77. The system of claim 76, wherein at least two generations of the first database are stored in globally accessible space, and at least one generation of the first database is stored remotely.

78. The system of claim 77, wherein two generations of the first database are stored locally.

79. The system of claim 78, wherein the analysis search can optionally query the first database stored globally or can copy the first database to a local node.

80. The system of claim 79, wherein the first database is transferred from globally accessible space using remote copy.

81. The system of claim 79, wherein the first database is transferred from globally accessible space using GridFTP.

82. The system of claim 65, wherein the periodic search is performed on a dedicated node called a periodic search node.

83. The system of claim 82, wherein the second database node comprises a storage means large enough to store at least two generations of the first database.

84. The system of claim 83, wherein the periodic search node utilizes an updating scheme, wherein the updating scheme allows updating of the first database and analysis searching of the database without interference.